TECHNICAL INNOVATION

A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression

Angela C. Burnett, Jeremiah Anderson, Kenneth J. Davidson, Kim S. Ely, Julien Lamour, Qianyu Li, Bailey D. Morrison, Dedi Yang, Alistair Rogers and Shawn P. Serbin*

Terrestrial Ecosystem Science and Technology Group, Environmental and Climate Sciences Department, Brookhaven National Laboratory, Upton, NY 11973, USA

* Correspondence: sserbin@bnl.gov

Abstract

Partial least squares regression (PLSR) modelling is a statistical technique for correlating datasets, and involves the fitting of a linear regression between two matrices. One application of PLSR enables leaf traits to be estimated from hyperspectral optical reflectance data, facilitating rapid, high-throughput, non-destructive plant phenotyping. This technique is of interest and importance in a wide range of contexts including crop breeding and ecosystem monitoring. The lack of a consensus in the literature on how to perform PLSR means that interpreting model results can be challenging, applying existing models to novel datasets can be impossible, and unknown or undisclosed assumptions can lead to incorrect or spurious predictions. We address this lack of consensus by proposing best practices for using PLSR to predict plant traits from leaf-level hyperspectral data, including a discussion of when PLSR is applicable, and recommendations for data collection. We provide a tutorial to demonstrate how to develop a PLSR model, in the form of an R script accompanying this manuscript. This practical guide will assist all those interpreting and using PLSR models to predict leaf traits from spectral data, and advocates for a unified approach to using PLSR for predicting traits from spectra in the plant sciences.

Keywords: Hyperspectral reflectance, leaf traits, LMA, modelling, plant traits, PLSR, spectra, spectroradiometer, spectroscopy.

Introduction

Plant leaf traits are the physiological, morphological, or biochemical characteristics of plants measured on individual leaves (Violle *et al.*, 2007). Leaf traits play an important role in plant resource acquisition and allocation, with broad impacts on primary production, community assembly, and plant responses to climate change (Reich *et al.*, 1997; Wright *et al.* 2004; Myers–Smith *et al.*, 2019; Kattge *et al.*, 2020; Thomas *et al.*, 2020). The use of leaf traits in ecological and evolutionary research has seen a dramatic increase in recent years, improving our understanding of vegetation from the scale of the individual plant to that of ecosystems (Woodward and Diament, 1991; Violle *et al.*, 2007; Bjorkman *et al.*, 2018). Furthermore, measuring leaf traits is a critical element of plant phenotyping, including phenotyping for crop breeding and precision agriculture

(Reynolds and Langridge, 2016). However, progress is hindered by the expensive and time-consuming acquisition of traits from *in situ* leaf harvesting and subsequent laboratory analysis (Cornelissen *et al.*, 2003). Laboratory processing of structural traits such as leaf mass per unit area (LMA) can take days to weeks to complete, while processing leaf samples for biochemical and physiological traits may take weeks to months and is also costly, requiring specialist equipment and expertise. To fully extend the potential of trait-based studies, robust and non-destructive methods for effectively identifying leaf traits are critically needed (Myers-Smith *et al.*, 2019).

Using leaf reflectance spectra to predict leaf traits is a promising alternative to traditional measurements, as reflectance can be quickly and efficiently measured using spectroradiometers. A leaf reflectance spectrum results from the light–absorbing and scattering properties of the leaf, and is highly modified by the concentration of light-absorbing compounds (photosynthetic pigments, water, cellulose, lignin, starch, protein, and macronutrients) and the internal structure of leaves which determine light scattering (Kumar *et al.*, 2002; Asner, 2008; Serbin and Townsend, 2020). As leaf structure and chemical composition vary with species, age, environment, and stress, these light-absorbing and scattering properties change in concert (Yang *et al.*, 2016; Wu *et al.*, 2017).

Typically, a reflectance spectrum contains a large array of variables, namely reflectance at different spectral wavelengths, ranging from several hundred to thousands depending on the spectral resolution and the range of measured wavelengths. These wavelengths typically include the visible (VIS, 380–700 nm), near infrared (NIR, 700–1100 nm), and short-wave infrared (SWIR, 1100–2500 nm) regions, as covered by a full-range spectroradiometer. While the spectrometer wavelength resolution will be likely to vary between detector regions, the spectral resolution is usually 3–5 nm in the VIS, and 6–12 in the NIR and SWIR bands, and the data are then interpolated to a 1 nm resolution. Given this large number of predictor variables, classic linear regression modelling cannot be used due to the problems of 'non-unique solutions' and 'overfitting' (Geladi and Kowalski, 1986; Wold *et al.*, 2001). Partial least squares regression (PLSR) was created to handle both the collinearity among predictors, in this case the different wavelengths of a reflectance spectrum, and the larger number of predictor variables than trait observations. PLSR projects the highly correlated predictor variables to a small number of latent variables and at the same time maximizes the correlation between the response and latent variables (Geladi and Kowalski, 1986; Wold *et al.*, 2001). This technique has been shown to be effective for handling spectral data (Serbin *et al.*, 2012, 2019; DuBois *et al.*, 2018; Wang *et al.*, 2020) and has been successfully implemented for many types of continuous traits across a variety of growth conditions, including glasshouses, farms, and natural ecosystems (Ely *et al.*, 2019; Meacham-Hensold *et al.*, 2019; Wu *et al.*, 2019). These traits include leaf structural traits, such as LMA (Serbin *et al.* 2019); biochemical traits, such as leaf

nitrogen, protein, sugars, starch, and lignin (Ely *et al.* 2019); and physiological traits, such as the maximum carboxylation rate of Rubisco ($V_{c,max}$) and the maximum electron transport rate (J_{max}) (Meacham-Hensold *et al.*, 2019; Wu *et al.*, 2019). These implementations of PLSR have largely improved our means for monitoring plant growth and plant physiological change under environmental stress, as well as the broad trait variation across vegetation types.

Previous work in the fields of chemometrics, medical sciences, and sociology has provided a detailed description of the mathematical aspects and advantages of PLSR over alternative approaches, therefore we will not address this here (Martens and Martens, 2000; Wold *et al.*, 2001; Krishnan *et al.*, 2011; Sawatsky *et al.*, 2015; Shmueli *et al.*, 2016). However, there is currently no clear guide on how to predict plant traits from hyperspectral data, and particularly how to apply the PLSR technique (Barnes *et al.*, 2017; Polley *et al.*, 2020; Villa *et al.*, 2020, Preprint). Whilst there are plenty of specific examples of PLSR being used to predict traits from spectra, we lack a general overview of PLSR for plant ecology.

Furthermore, there are a variety of approaches available for leaf trait and hyperspectral data collection, including choices of instrumentation, measurement procedures, and quality control, as well as different approaches for developing and validating a PLSR model, with steps including sample selection, handling imbalanced data and outliers, determining the number of components, and model validation. There is also a wide choice of PLSR software, such as 'pls' (R software environment), 'plsregress' (Matlab), 'sklearn' (Python), 'PLS procedure' in SAS, Statistica, and SPSS; furthermore, there are many different outputs, including figures and indices of performances. Collectively, this overwhelming variety of potential pathways for analysis is hindering the development of a coherent approach that can enable clear comprehension and comparison of results and the sharing and application of PLSR models by the community. Finally, for the researcher hoping to begin applying the PLSR technique to new questions, the absence of an established best-practice method leaves the uninitiated user unsure of how to begin working with this powerful technique.

Here we present a best-practice 'hands-on' guide to perform PLSR modelling for estimating plant traits from leaf-level hyperspectral data, and provide clear guidance on accuracy assessment and model uncertainty. We explain each part of the PLSR workflow from data collection to model building, validation, and application (Fig. 1), with key definitions provided in Box 1. We first examine the coordinated data collection of leaf traits and leaf spectra and suggest best practices. We then provide a tutorial guide through an example of a PLSR modelling scenario accompanied by a detailed, open-source R script. Finally, we discuss common pitfalls when using PLSR and tips for applying PLSR models to novel data. We therefore present a clear and practical guide for readers, reviewers, and researchers; advance knowledge of PLSR and its application; and provide a toolkit for performing one's own best-practice PLSR.
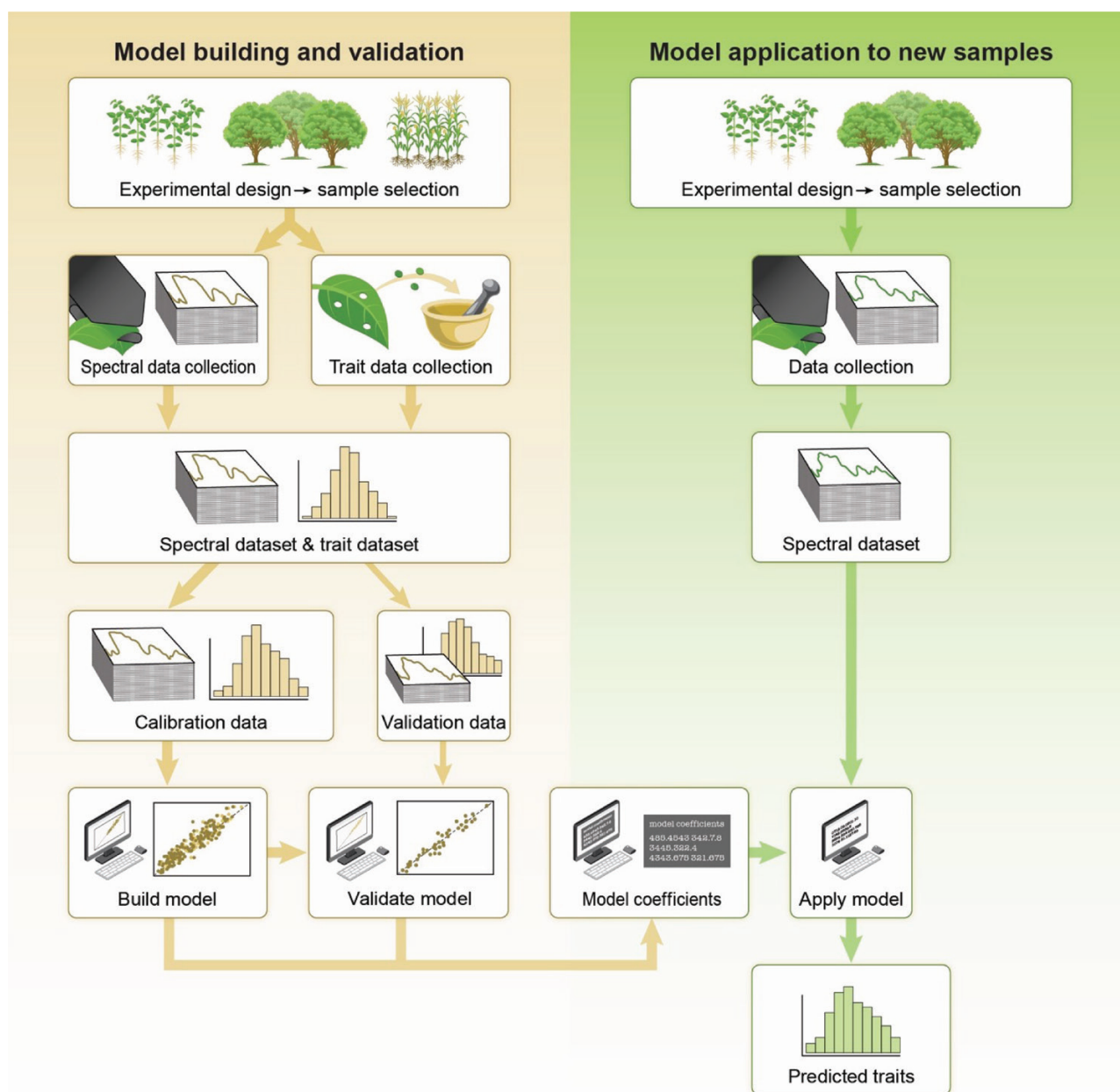
Fig. 1. Use of partial least squares regression (PLSR) for predicting leaf traits from leaf spectra. Left-hand side: the model is built and validated using a range of species, leaf ages, or experimental treatments to fill the desired trait space (in this example, we show a range of species). Spectra and traits are measured for each leaf of interest. The resulting dataset is split into calibration data and a smaller subset of validation data (Step 7 in the example R script), and used to build and validate a PLSR model as detailed in the text and accompanying R script. Once the dataset is split, the optimal number of components (Step 10 in the example R script) are determined, and the calibration model is developed (Step 11) and validated (Steps 11 and 12). In addition, an further uncertainty permutation analysis (Step 14) can be conducted to quantify the uncertainty in trait estimation. Right-hand side: new samples are selected from within the trait space used to build the model (in this example, the species range). Spectral data are collected for each leaf of interest. PLSR is performed using the model coefficients from the validated model in order to predict the trait of interest for each new spectrum.

Recommendations for collecting leaf spectra and trait data

In order to build a PLSR model with the best possible predictive power, data collection must be undertaken with the goal of assembling a dataset that captures the potential variation in the plant traits of interest. This involves sampling across the broadest possible trait space and using a spectral collection method that will accurately capture the trait; for some traits, timing and temperature are also factors to consider as outlined below. Following Violle *et al.* (2007), we

Box 1.   Definitions

Modelling

- Calibration: the process of training the model, using the calibration dataset. In models which are validated by cross-validation, typically 80% of the data are used to generate the model following a random stratified split. In models which are validated independently, 100% of the original data are used to generate the model.
- Collinearity: the case in which predictor variables are highly correlated, or not independent.
- Jackknife: a statistical resampling technique in which one sample is left out and the population parameter of interest is recalculated. The resulting mean of the resamples is the jackknife parameter estimate.
- PLSR (partial least squares regression): a regression technique for building a statistical model relating matrices of *x* and *y* variables.
- PRESS: predicted residual error sum of squares. The sum of squares of the prediction residuals for the observations used to validate the model; this provides a summary of model performance.
- Regression coefficients: the final linear regression coefficient values used to estimate the dependent variable *Y* from the elements of the predictor matrix *X* (spectral data). One of the core outputs of PLSR modelling is the regression coefficient vector containing the values to multiply each wavelength by (then add together plus the intercept) to estimate the dependent trait variable *Y*.
- Residual: the difference between the observed value and the predicted value.
- Trait model: a set of model coefficients that will allow the prediction of a leaf trait from a leaf reflectance spectrum.
- Validation: the process of testing the model. Cross-validation uses the split of the dataset not used in calibration, typically the remaining 20%. Alternatively, independent validation may be performed using a second (independent) dataset.
- VIP (variable influence on projection): a measure of the importance of a given predictor variable on prediction of a response variable. A value >0.8 is typically used as a cut-off for a highly significant predictor variable.

Traits

- Leaf trait: empirically measured physiological (e.g. $V_{c,max}$), morphological (e.g. LMA), or biochemical (e.g. starch) characteristics of plants measured on individual leaves.
- Trait space: the range of possible values for a trait across experimental variables such as species, leaf age, drought, nutrient treatments, or genetic variation.

Spectroscopy

- Hyperspectral data: spectral data collected over many contiguous narrow spectral bands (e.g. 1 nm), over multiple spectral regions (e.g. visible to short-wave infrared).
- Leaf reflectance spectra: a leaf reflectance spectrum describes the light-absorbing and scattering properties of the leaf.
- Spectral noise: variation in signal response resulting from limitations of accuracy and precision of the instrument used to collect spectral data.

define traits as empirically measured structural, biochemical, or physiological properties of leaves; this excludes properties such as the normalized difference vegetation index (NDVI) or output from a SPAD chlorophyll meter, since it is not meaningful to use leaf reflectance to predict properties that are themselves derived from reflectance within the same wavelength range.

Collection of leaf trait data

Non-linearities in plant traits mean that extrapolation of prediction beyond the range of measured values cannot be considered reliable (Schweiger, 2020). Therefore, a measurement plan should be devised to fill the trait space in the calibration data that is anticipated from the combinations of experimental

variables under consideration, such as species, leaf age, drought, nutrient treatments, and genetic variation (Fig. 2). For the purpose of building a new trait model, this can be achieved by selecting species that are known to have diversity across the target traits, or by manipulating the environment or plants in a way that will result in diversity of the trait of interest.

The minimum number of paired spectra and measured trait samples required to build a robust model is somewhat dependent on the nature of the trait itself. Enough samples must be collected to provide calibration across the desired data range for predicting traits. Typically, models with good predictive capacity can be built using ~100 samples (Kleinebecker *et al.*, 2009; Ainsworth *et al.*, 2014; Barnes *et al.*, 2017; Girard *et al.*, 2020). However, successful prediction of some traits may require several hundred samples, especially if the data are highly clustered or skewed. Construction of PLSR predictive models with >200 samples is not uncommon (Dechant *et al.*, 2017; Yendrek *et al.*, 2017; Silva-Perez *et al.*, 2018)

Spectra and trait measurements should generally be made on the same area of leaf material. This principle, and the nature of the trait analysis, will dictate the order in which the spectra and trait measurements are made. Spectra should be collected prior to any destructive sampling of the leaf material. Special cases for which it is appropriate to measure the spectra after trait collection, or on analogous leaves, are discussed below. Given that spectral measurements represent the area in the leaf sensor field of view, the physical, chemical, or physiological traits of interest should also be calculated on an area basis.

Collection of leaf spectral data

A variety of approaches may be used to collect spectral data, including on fresh and dried leaf material. Here we focus on reflectance measurements of fresh leaves using a full range

(350–2500 nm) spectroradiometer with a leaf clip or contact probe attachment. The following section describes a basic protocol for measuring leaf-level spectroscopy; refer to Schweiger (2020) for consideration of different instrument types and further protocol details. While we focus on full-range examples, researchers will typically remove data below 450 nm or 500 nm and >2400 nm corresponding to increased sensor noise, and lower signal-to-noise ratio, and thus potentially adding erroneous information or noise into the PLSR modelling. Moreover, the approaches we present are not specific to this spectral range and the PLSR method is also applicable to datasets with a reduced spectral range, such as within the VIS to NIR (400–1000 nm) range. In these cases, however, depending on the trait, the PLSR model performance may be reduced because the information provided in the other spectral wavelengths is missing (e.g. Doughty *et al.*, 2011).

Most off the shelf spectrometer systems that are supplied with a leaf clip or plant probe will also have an internal, calibrated light source. If the researcher plans to use a custom light source, it is important to consider any issues related to light intensity and uniformity across the spectral wavelengths of interest, and adjust collection conditions appropriately (e.g. integration time). However, most spectrometer systems also provide a means for automatic integration that would also adapt to different light sources. This topic is beyond the scope of this work; we refer the reader to Jacquemoud and Ustin (2019) and expect that the data being used have been properly prepared for use with PLSR modelling following their recommendations.

Ideally, leaves should be measured *in situ*, or as soon as possible following removal from the plant. Temporary storage in a cool, dark place with appropriate humidity control will minimize metabolic changes in the leaf (Foley *et al.*, 2006; Pérez-Harguindeguy *et al.*, 2013). It is important to ensure that the
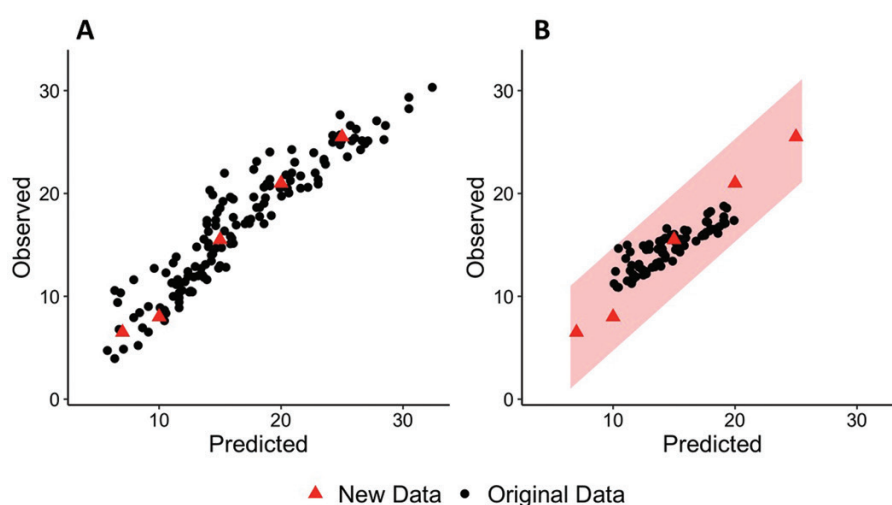

Fig. 2. Filling the trait space. PLSR models may be used to accurately predict values of traits from spectral data within the trait space filled by the original model. Black circles indicate the original data used to build the model. Red triangles indicate novel data to which the model is being applied. (A) Filled trait space. The model was built using data covering a broad trait space, which covers the range of new data points. New data can then be reliably predicted. (B) Unfilled trait space. The model cannot reliably predict the values of new data outside the trait space used to build the model. The trait space occupied by the new data is indicated by red shading.

leaf surface is dry before measuring and that all measurements are made on the same side of each leaf, usually the adaxial surface. Some sample types will require additional preparation before recording spectra, such as creating mats for narrow leaves or needles that will not cover the field of view of the sensor (Noda *et al.*, 2013; Kamoske *et al.*, 2021), or measurement of leaf temperature to allow scaling of temperature-sensitive traits (Silva-Perez *et al.*, 2018). It is also assumed that the researcher is familiar with the operation of their specific model of spectrometer and has followed the manufacturer's guidance on factory calibration, set-up, and warm-up procedures. For a complete treatment of this subject, see Jacquemond and Ustin (2019).

When the sample is ready, calibrate the spectroradiometer using a clean diffuse, 99% (or high reflectivity) reflectance standard. A common option is a LabSphere Spectralon diffuse reflectance standard (Lab Sphere, North Sutton, NH, USA). If the researcher would like to convert reflectance from relative reflectance to that corrected by the spectral response of the reflectance standard, it is important to record the serial number of the standard used so that a standard reference correction can be applied. This consists of multiplying each wavelength in the measured spectra by the reference standard calibration coefficient for that wavelength.

Once you have selected and recorded the standard, then take 3–5 measurements of the area to be harvested for trait measurement, avoiding irregular features such as midribs, prominent veins, insect damage or epiphyll cover. In certain plant types, the trait of interest may also influence the section of leaf chosen for analysis. For example, graminoids have well-characterized gradients of metabolites along the lamina that should be considered (Pick *et al.*, 2011). Care should be taken to minimize the time that the leaf is exposed to the light source to avoid damage from excessive heat. The bottom part of the leaf clip, or the leaf background if using a contact probe, should be optically dark (e.g. less than ~3% reflective) across all wavelengths being measured to minimize background reflectance. Use of a white background may result in the measurement of undesirable transflectance; this also limits the extension to canopy-scale modelling or synthesis with other datasets. After each measurement, check that the recorded spectrum is reasonable, with peak heights and positions as expected for the leaf type and water status (Serbin and Townsend, 2020). Poor quality measurements can arise from an incorrect reference measurement or improper placement of the leaf, including not filling the field of view, folding of the leaf, or not properly closing the leaf clip, leading to light leakage (Fig. 3). We recommend that users familiarize themselves with the typical spectral signature of the species or variant of interest to establish what characterizes good spectra for their study before commencing spectral data collection. The routine use of calibration standards for collected spectra is important for improving comparisons between spectral data collected by different operators and systems, and to facilitate interoperability of PLSR calibrations between datasets.
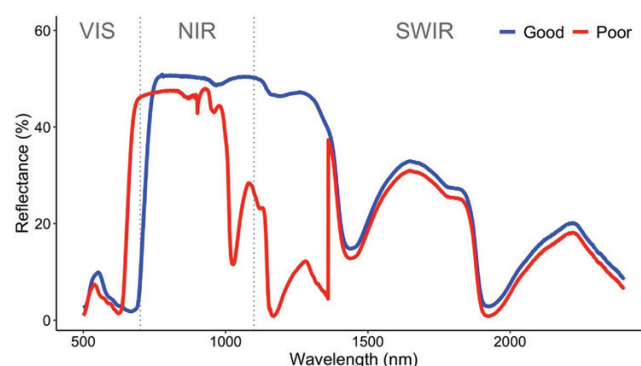

Fig. 3. Example leaf spectral data. The blue trace indicates a typical example of leaf spectral reflectance data; the red trace is an example of poor spectral data resulting from an error during measurement. These data are schematic, for illustrative purposes only. Spectral data shown here are from the range 500–2400 nm used in our PLSR example. Key regions of the spectrum are indicated with grey text and dashed lines (VIS, visible; NIR, near infrared; SWIR, short-wave infrared). The good example shows a peak in the visible spectrum relating to chlorophyll reflecting green light, then a strong increase in reflectance which should be between 35% and 80% (ideally between 40% and 60%) and should be the highest peak of the entire spectrum. This peak is the first of three consecutive peaks in the NIR and SWIR regions between ~750 nm and 1400 nm followed by a sharp decrease. After 1500 nm, there are two wide, final peaks, each lower than the preceding peak. The poor example (slightly offset from the good example to facilitate interpretation) shows a strong deviation from the expected pattern in the NIR and SWIR regions including the second NIR peak being higher than the first peak, and then sharp decreases and deviations between 1000 nm and 1500 nm, but the remaining regions show the expected pattern. Further examples of poor spectra could include the initial peak being <35% or >80% reflectance, or other deviations from the good example shown here.

To prepare the spectral data for model input, individual measurements should be reviewed and erroneous measurements removed (Fig. 3). Spectral artefacts resulting from detector overlap should be corrected, and the remaining spectra for each leaf sample should be averaged (Meacham-Hensold *et al.*, 2019) and labelled with a unique sample identifier. Similarly, the measured trait data should be assessed for quality, and erroneous or outlier measurements removed through the use of a standard technique such as boxplot or Cook's distance before commencing PLSR.

Special considerations for physiological traits

Physiological traits, such as $V_{c,max}$, have been derived from reflectance in a wide range of species and growth conditions (Serbin *et al.*, 2012; Yendrek *et al.*, 2017; Silva-Perez *et al.*, 2018; Meacham-Hensold *et al.*, 2019; Wu *et al.*, 2019). When modelling relationships between spectra and physiological traits, there are several additional considerations associated with the collection of reflectance data and expression of the physiological trait of interest. Measurement of leaf reflectance can result in damage to the leaf, particularly to sensitive physiological processes, so care must be taken to avoid overheating or over-illuminating the leaf (Serbin *et al.*, 2012). In addition, some physiological traits have a

strong temperature sensitivity, and changes in leaf temperature can also impact reflectance (Serbin, 2012; Khan *et al.*, 2021), thus it is important to either minimize heating of leaves during reflectance measurements or to pair leaf temperature measurements close in time with reflectance measurements. Certain physiological measurements can alter the leaf surface, for example when a sealant is used during gas exchange measurements to minimize leakage. The order of measurements will depend on the specific case, with the user considering the traits of interest to determine whether the reflectance measurement might adversely influence the physiological measurement, or vice versa. Whilst in an ideal scenario both measurements are made on the same region of the leaf, it is also possible to distribute measurements across the surface of a large leaf, on a different leaflet of a composite leaf, or even on an analogous sample of the same phenological stage and orientation on the plant. Great care should be taken to ensure that equivalent areas or leaves are measured.

The scaling of measurements to a reference temperature should also be considered. For example, the predictive power and accuracy of PLSR models have been shown to be greater when the model is built using maximum carboxylation rates that have been normalized to 25 °C ($V_{c,max.25}$) (Silva-Perez *et al.*, 2018) rather than the $V_{c,max}$ at measurement temperature, even though both approaches produce successful results (Serbin *et al.*, 2012). This may reflect the covariance between the biochemical and structural properties of the leaf that relate to the investment of resources in Rubisco including known (nitrogen, chlorophyll, and LMA) and unknown leaf traits that correlate with $V_{c,max.25}$ (Kattge *et al.*, 2009; Walker *et al.*, 2014; Croft *et al.*, 2017; Wu *et al.*, 2019). Given the role of environment, especially growth temperature, in determining the investment in Rubisco (Ali *et al.*, 2015; Kumarathunge *et al.*, 2019; Smith *et al.*, 2019) and the characteristics of other co-varying leaf traits, it is likely that the success of building PLSR models will be improved when the either the measurement temperature ($V_{c,max}$-based models) or chosen reference temperature ($V_{c,max.25}$-based models) is close to growth temperature. Leaf temperature should be recorded at the time of taking the spectral measurement if it is desired to scale $V_{c,max}$ to leaf temperature. Leaf temperature data are also required to use spectra to estimate the temperature response of $V_{c,max}$. In some cases, physiological parameters have been derived using species-specific kinetic constants (Silva-Perez *et al.*, 2018). We believe that this extra step is not necessary given the intended use of the spectra–trait approach as a screening tool and believe that use of standard kinetic constants (Bernacchi *et al.*, 2001) should be sufficient to develop effective PLSR models. Finally, we note that whilst $V_{c,max}$ changes with leaf temperature, effects of leaf temperature on the prediction of structural traits such as LMA are small (Khan *et al.*, 2021).

Special considerations for biochemical traits

Biochemical traits comprise another category of leaf traits that may be of particular interest for prediction from spectra. These traits include foliar protein, starch, and chlorophyll content. There is a long history of using reflectance spectroscopy to detect the concentration of chemical constituents in dried and fresh foliar material in the laboratory, using simple, stepwise, and PLSR methods (Curran *et al.*, 1992; Martin and Aber, 1997; Ustin *et al.*, 1998; Merzlyak *et al.*, 2003; Kokaly *et al.*, 2009; Ely *et al.*, 2019). All such models rely on the basic principle that the depth and magnitude of spectral absorption features associated with molecular bonds should vary in proportion to the concentration of that chemical (Curran, 1989; Kokaly, 1999). For example, N–H bonds, common in proteins, have vibrational absorptions at ~1980 nm and ~2130–2180 nm, with harmonic and overtone frequencies in the NIR and SWIR regions (Curran, 1989; Kokaly, 2001). Therefore, a model for proteins should rely heavily on the reflectance in these key wavebands (Curran, 1989). This should, in theory, make biochemical traits easier to model than structural or physiological traits, as the latter are the consequences of multiple chemical and structural constituents within the sample of interest, and thus direct correlative models are often confounded (Serbin and Townsend, 2020).

A number of considerations still need to be made in order to successfully construct a PLSR model for a biochemical trait. It is prudent to consider both the range and resolution of wavebands measured by the spectroradiometer used for data collection. The wavelength range should be such that the crucial wavebands are measured, and the resolution should be high enough to distinguish bond-specific absorption peaks from other peaks of similar bond types (Yoder and Pettigrew-Crosby, 1995). Furthermore, it is important to be aware that some chemical types may be masked by absorption associated with other chemicals with a common bond. For example, O–H bonds, the bond type in water, have strong absorption features at ~970, ~1200, ~1400, ~1450, and ~1940 nm (Curran, 1989). The O–H bond is also common in cellulose, starch, and lignin. Since most fresh plant material is highly water saturated, a model which leverages the O–H bond absorption features to predict compounds other than water may be confounded if the samples vary in their hydration. In addition, the strong dominance of water on the SWIR section of the electromagnetic spectrum may mask or obscure other subtle absorption features in the region (Curran, 1989). Drying of sample material will limit the influence of water's absorption features on spectra; however, this is a destructive step that may be impractical for paired spectra and biochemical sampling, which typically relies on flash-freezing samples for subsequent biochemical analysis.

The time of day of sampling is another consideration, especially for biochemical traits for which the foliar concentration will vary throughout the day as the plant mobilizes and redistributes certain types of biochemical compounds (Nozue and Maloof, 2006). For example, sucrose and starch content will be relatively low at dawn but will increase over the course of the day, peaking between solar noon and the end of the photoperiod, depending on the species and environmental

conditions (Sicher *et al.*, 1984; Hendrix and Huber, 1986). This time of day effect must be accounted for when planning spectral data collection, and can be leveraged to maximize differences in foliar biochemical concentration in order to fill the trait space (Fig. 2).

Tutorial for performing PLSR

Here we provide general guidelines and typical methods that can be applied to leaf-level PLSR spectra–trait modelling efforts, with regard to data preparation, model development and validation, and interpretation of results. We provide these guidelines in order to improve the robustness of the developed PLSR, but also, critically, to enable easier cross-comparison of published PLSR modelling results across datasets, teams, projects, and biomes. The detailed steps for building a PLSR model will vary for a number of reasons, including the trait of interest, wavelength range of the spectra (e.g. VIS to NIR versus VIS to SWIR), and the goal of the modelling effort; thus, it is not possible to provide a single generalized example to cover all applications. To illustrate our suggested approach, we provide the 'spectratrait' package (Supplementary Protocol S2 at *JXB* online), written in the open-source R statistical programming language, which contains examples that utilize publicly available leaf spectra and trait data (see Supplementary Table S1) provided through the Ecological Spectral Information System (EcoSIS; https://ecosis.org/). We provide a tutorial (Burnett *et al.*, 2020) to cover what is presented graphically in Fig. 1 through a GitHub repository (https://zenodo.org/record/4730995). The specific example R script summarized here is: 'spectra-trait_reseco_lma_plsr_example.R', with a full illustration of the results available online (https://rpubs.com/sserbin/736861). Users should begin with the README file available in the GitHub repository, which explains the package and illustrative vignettes provided. Before using the code on their own datasets, users should ensure that spectral and trait data have been curated in accordance with the recommendations provided above.

Data import and preparation

While specific to the examples here using EcoSIS datasets, data preparation and import into the chosen statistical modelling environment is a necessary early step (following data quality control). In the case provided here, we first define the required libraries and other options (Steps 1–3), including setting up output directories, prior to data import. Step 4 imports the example dataset from EcoSIS using the EcoSIS application programmer interface (API), using each dataset's API key with the 'get_ecosis_data.R' function. Steps 5 and 6 show an example of preparing data for fitting and removing missing or bad datasets. This includes the selection of the wavelengths to use in model

fitting, in this case 500–2400 nm. Individual researchers should determine the level of data curation and preparation necessary for their specific case, as well as the appropriate wavelengths to use.

Data transformation

Although PLSR does not rely on 'hard' distribution assumptions (Goodhue *et al.*, 2012; Sawatsky *et al.*, 2015), better results are obtained when the trait distribution is close to normal or is not significantly skewed. Several methods can be used to inspect the normality of the trait data, including histograms, Q–Q plots, and normality test methods; we use a histogram for illustration in the example code (Step 7). If the histogram does not follow a symmetrical distribution, corresponding transformations should be applied to improve its symmetry before commencing PLSR. For example, a log transformation can be used on data with a long tail on the histogram (i.e. fewer measurements at large trait values). If the value zero occurs in a variable, the fourth root transformation is a good alternative to the logarithm (Wold *et al.*, 2001). While we suggest transforming the trait data if needed, we do not recommend transforming leaf-level spectral data because the reflectance of all wavelengths is collected in the same percentage unit (i.e. 1–100%) and the use of a leaf clip minimizes environmental interference, meaning that transformation to improve the data distribution is not required.

Data splitting for PLSR calibration and validation

Early PLSR spectra–trait modelling efforts relied solely on internal cross-validation to evaluate model performance (Townsend *et al.*, 2003). More recently, we have instead advocated splitting the full dataset into a model training (calibration) dataset and an out-of-sample validation dataset to provide a more robust and accurate evaluation of prediction, particularly for large datasets (e.g. Serbin *et al.*, 2014). The commonly used cut-offs for calibration versus validation vary between 60% and 80% (Meacham–Hensold *et al.*, 2019; Serbin *et al.*, 2019; Wu *et al.*, 2019). In our example, Step 7 illustrates one way of splitting the data. We use 80% for PLSR calibration in our demonstration (i.e. 80% of the spectra–trait data pairs). Figure S1 in Supplementary Protocol S1 shows the resulting calibration and validation trait distributions based on the split criterion provided. While it is not possible to define all the potential experimental variables that an investigator may use to guide data splitting, such as treatment, species, and spatial range, it is useful to consider one or more key factors that help to define the variation in datasets. For example, 'species' is accounted for in our main example, while other examples provided show other variables, such as the NEON (National Science Foundation's National Ecological Observatory Network) example which partitions the data based on an additional variable 'domain' corresponding to specific NEON

domains (R script: spectra-trait_neon_lma_plsr_example.R). During the splitting procedure, the 80/20%, for example, is applied to each bin variable, such as species. Similarly, experimental treatments should also be carefully considered if the numbers of collections under each treatment or species are significantly different from each other. The ultimate goal of these considerations is to prevent fewer observed species/domains/treatments from being ignored or under-represented in PLSR calibration.

Once the data have been split for conducting PLSR, formatting may be needed in order to use the data with the chosen pls package. For example, we illustrate this formatting requirement in Step 8 where we define the PLSR dataset as required by the package. We also show the calibration and validation spectra data side-by-side in Fig. S2 in Supplementary Protocol S1.

Data permutation to determine the optimal number of components

Selecting an optimal number of components (NoC) is critical for the performance of PLSR. Selecting an erroneously small NoC can result in underfitting of the spectra data and poor model performance. On the other hand, overfitting the model by selecting too many components can yield a model with erroneously high training statistics but will more than likely lead to low model performance, particularly when applied to new observations. This is because overfitting will tend to fit the model to spurious noise in the higher order components. If there is a lot of error or unexplained variance in the data, or if there is the need to link numerous different observations or data types together (Wolter *et al.*, 2008), modelling can require tens of components. However, our experience is that 20 components is a typical maximum for fitting basic, leaf-level spectra–trait models (Fig. S3 in Supplementary Protocol S1). To optimize the process of selecting the NoC, we advocate data permutation to identify the optimal NoC and provide the ability to statistically select the smallest necessary to optimize prediction (e.g. Serbin *et al.*, 2014; Ely *et al.*, 2019). In our examples, we provide three methods that illustrate ways to permute and select the optimal NoC.

To perform a data permutation, a sufficient number of PLSR models (20 in our example) need to be built for each of a series of NoC. The averaged predictive performance of these PLSR models is used to determine the optimal NoC. For each model, a subset of samples is randomly selected from the calibration data to train the model, with the rest used for model validation. Here we refer to these as 'jackknife calibration' and 'jackknife validation' samples to differentiate them from the PLSR calibration and validation subsets. Step 10 shows one of the ways to conduct the data permutation. In our example, we used 70% of the calibration subset for jackknife calibration and 30% for jackknife validation, an approach similar to Serbin *et al.* (2014) or Couture *et al.* (2016). Following this, the predicted residual error sum of squares (PRESS) can be calculated for each NoC across all PLSR model ensemble members (Fig. S3 in Supplementary Protocol S1). We do not recommend low data splits for jackknife validation, such as 10% (e.g. Asner *et al.*, 2014), as this may lead to overfitting due to an inability to statistically differentiate models with larger numbers of components, particularly with very large calibration datasets.

Typically, the PRESS plot has a V-shape (decrease then increase) from a low to high NoC, and the optimal NoC is indicated by the minimum PRESS (e.g. vertical blue line in Fig. S3 in Supplementary Protocol S1). However, this V-shape is not universal. Patterns such as a slowly decreasing PRESS at larger NoC can be obtained when a high variance is present in the spectral data. Despite the decreasing PRESS, 'overfitting' may already exist in PLSR with a large NoC, and diminishes the applicability of the model to new datasets. In this case, a threshold can be set to stop the NoC selection at which the change of PRESS becomes insignificant, for example if the change of PRESS from n to $n+1$ NoC is <5%, or a t-test on PRESS does not show significant differences between n and $n+1$ ($P>0.05$). In the example code, we demonstrated this with a t-test. Starting from the first NoC, the t-test is conducted between any two consecutive NoC, and stops when the P-value is >0.05.

Final PLSR model calibration and validation

Following the selection of the optimal PLSR NoC, a final model should be calibrated using the previously generated calibration data subset. For more detail, see the information in the R *plsr.out* object for PLSR model features; details for parameters can be found in Mevik *et al.* (2015). In our example, this is shown in Step 11 where we refit the model with the calibration (training) data to define the final model coefficients.

Assessment of the final calibration model can be done in two ways. First, leave-one-out cross-validation (LOO-CV), sequential or 'Venetian blinds' cross-validation can be used to internally assess the model training fit. However, this will probably over-represent the actual model performance. A more appropriate assessment is to utilize the out-of-sample validation dataset as validation of the model performance. This approach is generally effective when both the calibration and validation data are representative of the trait space. For example, we show the assessment of the LMA model using our withheld validation data across increasing PLSR model components up to the final optimal number, 11 (Supplementaty Fig. S4 in the GitHub repository). We show both how the root mean square error of prediction (RMSEP) decreases and how the coefficient of determination (R^2) increases with successive components to a final RMSEP of 11.7 g m^{-2} and R^2 of 0.86. It is, however, important to note that overestimation of model performance is still a potential issue in our proposed approach as the calibration and validation subsets are not fully independent.

Additional steps may be considered to reaffirm the perform-ance of the final PLSR. For example, an extra independent dataset can be used to avoid model overestimation and also test its generalizability to new observations. While this could seem unnecessary, often the overall goal of PLSR modelling is to develop models that can be used to estimate traits for new samples (e.g. Serbin *et al.*, 2019) so this additional step may be important in some cases.

As we show in Fig. 4, there are various metrics to assess model performance. These include the R^2, RMSEP, and the percent RMSEP (RMSEP as a percentage of the trait range). In Step 12 of our example, we further illustrate a way to evaluate model fit through both calibration and validation fit statistics and assessment of residuals. For example, Fig. S5 in Supplementary Protocol S1 provides a side-by-side assessment of the LMA model performance for the calibration and valid-ation of this model using the withheld dataset. The calibration and validation residual plots are provided to help visually assess any significant bias or skewness in the model fit. For example, when reviewing the residuals, if a strong non-linear trend is found it may indicate that data transformation (e.g. square root

Fig. 4. Selected PLSR model outputs. (A) Jackknife regression coefficients showing the model uncertainty based on the calibration data. The magnitude of the coefficients in the positive or negative direction provides an assessment of the strength of the contribution of wavelengths or wavelength regions to the prediction of the leaf trait, while coefficients close to zero indicate a very low contribution of that wavelength to prediction. PLSR model coefficients are also useful to diagnose for contributions of other constituents (Dechant *et al.*,2017) based on known absorption features (Curran, 1989). (B) Variable influence on projection (VIP) indicates the importance of a specific wavelength for predicting the trait of interest, here LMA. A VIP >0.8 typically indicates a variable of high importance to model prediction (Wold *et al.*, 2001). Together, these graphs help the user to understand a PLSR model, critical wavelength regions for prediction, and possible covariance with other traits based on known absorption features (e.g. Serbin *et al.*, 2014) or by plotting coefficients and VIPs of several traits together to explore similarity (e.g. Ely *et al.*, 2019).

transform, Serbin *et al.*, 2019) may be required. In this specific example, we see that the model R^2 is similar for the calibration and validation results, while the RMSEP increased slightly for the validation data, from 12.84 g m^{-2} to 13.68 g m^{-2}.

An additional step that is often useful for assessing PLSR models is to explore the final model coefficients and vari-able influences on projection (VIPs; Fig. S6 in Supplementary Protocol S1), and this is shown in Step 13 in the tutorial ex-ample. In this case, we plot the final 11 component PLSR re-gression coefficients and the VIPs across the same wavelength range. Here we see that the areas of highest contribution to the estimation of LMA in this dataset tend to occur in the vis-ible wavelengths, followed by portions of the SWIR regions, with a slightly lower contribution from the NIR bands. We can also see that the resulting coefficient plot shows reason-ably smooth values across most of the spectrum, with some areas that have a higher frequency variation related to regions of spectrometer detector overlap where the signal-to-noise is often lower, or other features that may be related to measure-ment artefacts. The coefficient plot can also be used to iden-tify spectral features related to specific constituents. Known spectral peaks can be retrieved and used to give a possible explanation of why the model successfully predicts the trait of interest (Dechant *et al.*, 2017). Typically, a high degree of high-frequency variation in a coefficient plot may indicate an overfit model (Asner *et al.*, 2014).

PLSR model uncertainty analysis

In our example, we have illustrated how to prepare, define op-timal components, fit, and evaluate a PLSR model using cross-validation, permutation analysis, and independent validation. However, it may be of interest for researchers to understand the inherent uncertainties in their model when predicting new values based on leaf spectra. This can be achieved by deriving model predictive uncertainty via a permutation ana-lysis (Serbin *et al.*, 2014; Couture *et al.*, 2016; Ely *et al.*, 2019; Wu *et al.*, 2019). Most commonly this is done by: (i) generating PLSR ensembles through permutation (e.g. jackknife or boot-strapping); (ii) collating these results into a matrix of possible models; (iii) applying the ensembles to new observations to derive a distribution of fitted values for each observation; and then (iv) summarizing the fit uncertainty given the variance around each estimated value. In addition, once the model en-sembles are generated, they may be used with the original final model to estimate new trait values and the uncertainties around these estimates employing the permuted model coeffi-cients (e.g. Serbin *et al.*, 2019).

There are various ways to conduct permutation analysis (Singh *et al.*, 2015; Couture *et al.*, 2016; Schweiger *et al.*, 2018) but, in this case, we followed a basic jackknifing approach. Step 14 in our example illustrates the internal jackknife feature of the PLS package in R, which we used to develop the PLSR model ensembles for the uncertainty assessment. The resulting
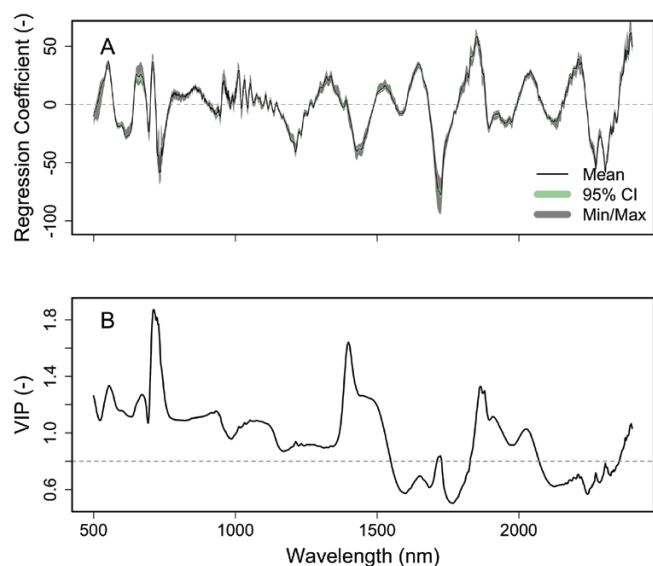
jackknife coefficients are shown in Fig. S7 in Supplementary Protocol S1. As expected, the general shape of the jackknife coefficients across the spectral range is similar to the full model (Fig. S6 in Supplementary Protocol S), but also shows the variance in coefficients related to the resampling of the full calibration data that are used to quantify and represent the uncertainty of the PLSR model. Using the full model and jackknife coefficients, we then provide the mean estimate as well as both the 95% confidence and prediction intervals of the estimated LMA values in the validation dataset. We show the estimated uncertainty in Fig. S8 in in Supplementary Protocol S1, which demonstrates the standard observed versus predicted presentation of the results, with R^2, RMSEP, and %RMSEP, as well as the prediction interval around each validation data point. This is similar to that provided by previous studies (Couture *et al.*, 2016; Ely *et al.*, 2019; Wu *et al.*, 2019) and is a recommended way to present overall model predictive performance in a scientific paper.

Final steps

After fitting the model and estimating the inherent uncertainty, the last step is to output or save the model, and fit statistics, diagnostics, final model coefficients, and permutation coefficients. Importantly, the model coefficients and permutation coefficients represent the resulting model and the resulting model with uncertainties, respectively. These coefficients are the main features used to estimate traits from new leaf spectral observations. We illustrate saving PLSR results in Steps 15 and 16 in the example.

Common pitfalls using PLSR when predicting plant traits

In this article we have highlighted the uses and applications of PLSR modelling in the context of estimating leaf traits with high-resolution leaf-level spectroscopy data. While this method has proved useful in a range of applications from forestry, agriculture, plant biology, and remote sensing, it is not the only method available—and is not always the best method—for building links between plant functional properties and remote sensing signatures. For example, PLSR can suffer from a strong influence from outliers, even if the fraction of erroneous or otherwise questionable data is small relative to the total number of observations (Wold *et al.*, 2001). Moreover, as stated here and elsewhere (Schweiger, 2020), PLSR, like all empirical methods, requires that the training dataset encompass the range of expected values for a trait to which the model will be applied otherwise it will probably underperform at the extremes of the distribution (Fig. 2). As such, it is recommended that researchers who wish to develop models for operational applications, or across diverse vegetation, focus on building large datasets that cover the trait space they expect

(e.g. Serbin *et al.*, 2019). This, however, can present a significant challenge as the cost and effort to develop a model for any one location or experiment can be large when factoring in logistical costs for data collection. Therefore, we recommend that datasets such as those available from EcoSIS be leveraged whenever possible to expand the training dataset to encompass more expected variation in spectra and trait observations. Past efforts suggest that this is possible (e.g. Serbin *et al.*, 2019), but additional research into the ease of combining trait and spectral collection methods, instruments, and other conditions is still required to fully understand the generality of PLSR approaches across large and diverse datasets. Furthermore, other issues can impact the performance of PLSR modelling such as noisy spectral data, particularly in the regions of known absorption features of certain traits such as the SWIR region; this is common with the use of some fore optics such as integrating spheres (e.g. Yang *et al.*, 2016).

Another pitfall of PLSR modelling is the potential to overfit the training data. PLSR modelling depends on the decomposition of the original data into a number of orthogonal axes equal to the number of different components selected in the modelling step. This is a critical step to develop the fewest number of different rotations or orthogonal axes needed to optimally model the dependent variable(s) [Y] using the predictor matrix [X]. This makes this method more suitable than linear regressions by reducing the number of variables used in the model. However, in this step, it is possible to overfit the PLSR model if too many components are chosen. In that case, the model will start to use the noise as if it were structural information to explain the variation in Y. This effect is similar to trying to fit a linear model with as many parameters as the number of observations. In that case, the model will describe exactly the observation but the parameters will not be significant. When this happens, it is possible to overtrain the model with a very large number of components. The more components included in the model, the more likely it is that spurious noise is included. For most plant traits, <20 should be all that is required. As a result of the overfitting step, the training may suggest a performing model but its evaluation on the validation dataset gives a lower prediction accuracy. We recommend that researchers follow a strict component selection protocol (see provided example scripts) and explore outlier removal prior to model training. This will help ensure models are not overfit.

Reporting and applying PLSR models to new datasets

Finally, we recommend practices for reporting PLSR models that aid interpretation of results and model reuse. PLSR is a robust method for inferring leaf properties from spectroscopic measurements and for reducing the risk of overfitting. However, the way in which PLSR is carried out and the choice of the number of components is of paramount importance.

Therefore, the chosen method should always be carefully described, and the number of components should be as parsimonious as possible so other researchers can use the model with new data while having the same predictive power. The use of a validation dataset to evaluate PLSR is also very important for reducing the risk of overestimating the quality of the model. The RMSEP and R^2 are the recommended indicators of performance and should both be reported. %RMSEP is useful for facilitating comparisons between traits for which the data ranges span different scales. Figure 5 provides an example of a figure and legend to report the performance of a PLSR model. Model reporting should also include the number of points of the validation dataset (*n*) and the NoC. Other metrics and figures could also be added if needed, such as the mean absolute error (MAE) which is less sensitive to outliers than the RMSEP. The histogram of the residuals can also be interesting to report as it contains information on the outliers and under- versus overprediction. (The code to produce those figures is provided in the GitHub repository.) However, we suggest always using the RMSEP and R^2 as a minimal basis for all PLSR descriptions to allow direct comparison between models.

In order for other researchers to utilize an existing PLSR model with new datasets, the coefficients of the PLSR model, the intercept, and any transformations of the variable should also be published. For example, GitHub (e.g. https://github.com/serbinsh/SSerbin_etal_2019_NewPhytologist/releases/

Fig. 5. Example observed versus predicted plot with prediction uncertainty. Final validation plot resulting from model validation using the validation dataset. The relationship between observed (measured) and predicted trait values (from PLSR) is shown, with the dashed black line indicating a 1:1 relationship. The grey error bars are the 95% prediction interval around each estimated LMA based on the jackknife analysis (Fig. 4). Validation R^2, RMSEP (root mean squared error of prediction), and %RMSEP (RMSEP as a percentage of the range of observed trait values) are shown.

tag/1.2.1) or the Ecological Spectral Model Library (EcoSML.org) model aggregator portal can be used to provide access to the final PLSR models.

With the information reported above, potential users can determine if an existing model is suitable for their research needs. However, special consideration is needed to determine if the PLSR model can be applied in their environment and for their specific samples. First, it is always good practice to assess the quality of the model using at least 20 new samples. Second, it is important to note that the models should be applied using the same protocol and the same type of equipment as used in the original model development. An attractive use of leaf-level PLSR models would be to apply them on a larger scale using other sensors such as multispectral or hyperspectral cameras that can be deployed in unoccupied aerial systems, aircraft, or satellite platforms. Such an aim is not straightforward, mainly because the 3D position of the leaves and their light environment are not uniform between samples as they are for proximal spectral data collection, and this means that these aspects must be accounted for when developing a canopy model. In addition, sensor characteristics such as the bandwidth and number of wavebands may vary between instruments. As a consequence, the application of a PLSR model built at the leaf scale to canopy spectral data can lead to a significant bias (Al Makdessi *et al.*, 2019). Instead, the structure of a canopy and sensor characteristics must be accounted for when building a PLSR model. In practice, this means that such a model should be built and applied at the canopy scale, rather than building a leaf-level model and applying it at the canopy scale (e.g. Burnett *et al.*, 2021). A detailed discussion of these points is outside the scope of this manuscript, although for the interested reader we provide an example of a canopy-scale application of our approach using a dedicated model (see Supplementary Table S1). Lastly, the PLSR model should not be applied on new samples outside the range of the original dataset used to calibrate the PLSR model (Fig. 2B).

In conclusion, this practical guide to PLSR models will enable researchers to better understand and use this powerful technique for predicting leaf traits from leaf-level spectral data—both for building new spectra–trait models and for applying existing models to new spectral data. The detailed tutorial examples can be used as a guide to achieving optimal results when building new models, which will, in turn, facilitate a higher degree of confidence in applying these or other published models to new spectral datasets, and leveraging existing datasets to examine leaf traits.
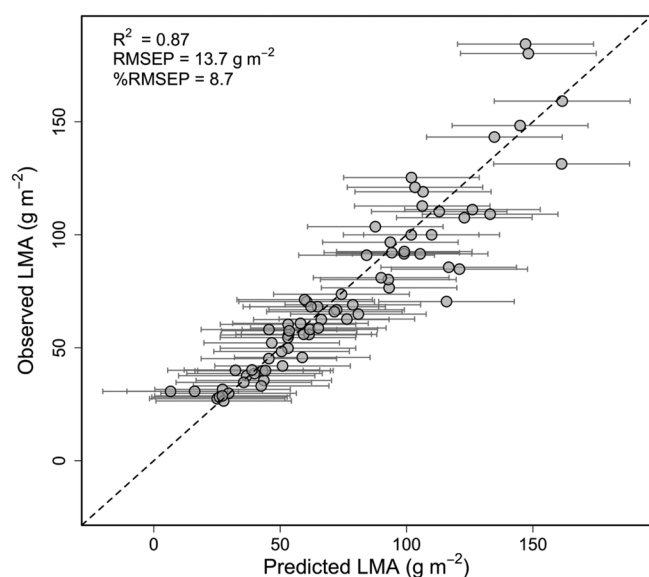
Supplementary data

The following supplementary data are available at *JXB* online.

Table 1. Details of the R scripts provided in our GitHub repository.

Protocol S1. Illustration of our PLSR tutorial.

Protocol S2. A compressed zip file containing the 'spectratrait' R package (spectratrait–v1.0.4.zip); PDF file detailing the contents of the zipped supplementary R package.

Supplementary figures referenced in the manuscript are included in the vignette and as outputs of the R script.

Acknowledgements

Author contributions

ACB, AR, and SPS: conceptualization; ACB: manuscript preparation; SPS: development of the R script with significant contributions from JA and JL, and additional contributions from all authors. All authors contributed to the writing of the manuscript.

Data availability

All illustrative examples in this study use published, open–source data which are available online through the Ecological Spectral Information System (EcoSIS, https://ecosis.org/) as outlined within the manuscript and Supplementary Table S1. The R script developed to accompany this manuscript, along with a readme file and illustrative vignettes, are located on Zenodo (https://zenodo.org/record/4730995). A link to our repository has been uploaded as a protocol on protocols.io doi: dx.doi.org/10.17504/protocols.io.bmhek33e (Burnett *et al.*, 2020).

References

Ainsworth EA, Serbin SP, Skoneczka JA, Townsend PA. 2014. Using leaf optical properties to detect ozone effects on foliar biochemistry. Photosynthesis Research **119**, 65–76.

Al Makdessi N, Ecarnot M, Roumet P, Rabatel G. 2019. A spectral correction method for multi-scattering effects in close range hyperspectral imagery of vegetation scenes: application to nitrogen content assessment in wheat. Precision Agriculture **20**, 237–259.

Ali AA, Xu C, Rogers A, et al. 2015. Global-scale environmental control of plant photosynthetic capacity. Ecological Applications **25**, 2349–2365.

Asner G. 2008. Hyperspectral remote sensing of canopy chemistry, physiology, and biodiversity in tropical rainforests. In: Kalacska M, Sanchez-Azofeifa GA, eds. Hyperspectral remote sensing of tropical and sub-tropical forests. Boca Raton, FL: CRC Press, 261–296.

Asner GP, Martin RE, Carranza-Jiménez L, Sinca F, Tupayachi R, Anderson CB, Martinez P. 2014. Functional and biological diversity of foliar spectra in tree canopies throughout the Andes to Amazon region. New Phytologist **204**, 127–139.

Barnes ML, Breshears DD, Law DJ, van Leeuwen WJD, Monson RK, Fojtik AC, Barron-Gafford GA, Moore DJP. 2017. Beyond greenness: detecting temporal changes in photosynthetic capacity with hyperspectral reflectance data. PLoS One **12**, e0189539.

Bernacchi CJ, Singsaas EL, Pimentel C, Portis AR, Long SP. 2001. Improved temperature response functions for models of Rubisco-limited photosynthesis. Plant, Cell & Environment **24**, 253–259.

Bjorkman AD, Myers-Smith IH, Elmendorf SC, et al. 2018. Plant functional trait change across a warming tundra biome. Nature **562**, 57–62.

Burnett AC, Anderson J, Davidson KJ, Ely KS, Lamour J, Li Q, Morrison B, Yang D, Rogers A, Serbin SP. 2020. Example PLSR for predicting leaf traits from leaf spectra. protocols.io https://dx.doi.org/10.17504/protocols.io.bmhek33e

Burnett AC, Serbin SP, Rogers A. 2021. Source:sink imbalance detected with leaf and canopy-level spectroscopy in a field-grown crop. Plant, Cell & Environment doi: 10.1111/pce.14056.

Cornelissen JHC, Lavorel S, Garnier E, et al. 2003. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. Australian Journal of Botany **51**, 335–380.

Couture JJ, Singh A, Rubert-Nason KF, Serbin SP, Lindroth RL, Townsend PA. 2016. Spectroscopic determination of ecologically relevant plant secondary metabolites. Methods in Ecology and Evolution **7**, 1402–1412.

Croft H, Chen JM, Luo X, Bartlett P, Chen B, Staebler RM. 2017. Leaf chlorophyll content as a proxy for leaf photosynthetic capacity. Global Change Biology **23**, 3513–3524.

Curran PJ. 1989. Remote sensing of foliar chemistry. Remote Sensing of Environment **30**, 271–278.

Curran PJ, Dungan JL, Macler BA, Plummer SE, Peterson DL. 1992. Reflectance spectroscopy of fresh whole leaves for the estimation of chemical concentration. Remote Sensing of Environment **39**, 153–166.

Dechant B, Cuntz M, Vohland M, Schulz E, Doktor D. 2017. Estimation of photosynthesis traits from leaf reflectance spectra: correlation to nitrogen content as the dominant mechanism. Remote Sensing of Environment **196**, 279–292.

Doughty CE, Asner GP, Martin RE. 2011. Predicting tropical plant physiology from leaf and canopy spectroscopy. Oecologia **165**, 289–299.

DuBois S, Desai AR, Singh A, et al. 2018. Using imaging spectroscopy to detect variation in terrestrial ecosystem productivity across a water-stressed landscape. Ecological Applications **28**, 1313–1324.

Ely KS, Burnett AC, Lieberman-Cribbin W, Serbin SP, Rogers A. 2019. Spectroscopy can predict key leaf traits associated with source–sink balance and carbon–nitrogen status. Journal of Experimental Botany **70**, 1789–1799.

Foley S, Rivard B, Sanchezazofeifa G, Calvo J. 2006. Foliar spectral properties following leaf clipping and implications for handling techniques. Remote Sensing of Environment **103**, 265–275.

Geladi P, Kowalski BR. 1986. Partial least-squares regression: a tutorial. Analytica Chimica Acta **185**, 1–17.

Girard A, Schweiger AK, Carteron A, Kalacska M, Laliberté E. 2020. Foliar spectra and traits of bog plants across nitrogen deposition gradients. Remote Sensing **12**, 2488.

Goodhue DL, Lewis W, Thompson R. 2012. Does PLS have advantages for small sample size or non-normal data? MIS Quarterly **36**, 981–1001.

Hendrix DL, Huber SC. 1986. Diurnal fluctuations in cotton leaf carbon export, carbohydrate content, and sucrose synthesizing enzymes. Plant Physiology **81**, 584–586.

Jacquemoud S, Ustin S. 2019. Leaf optical properties. Cambridge: Cambridge University Press.

Kamoske AG, Dahlin KM, Serbin SP, Stark SC. 2021. Leaf traits and canopy structure together explain canopy functional diversity: an airborne remote sensing approach. Ecological Applications **31**, e02230.

Kattge J, Bönisch G, Díaz S, et al. 2020. TRY plant trait database—enhanced coverage and open access. Global Change Biology **26**, 119–188.

Kattge J, Knorr W, Raddatz T, Wirth C. 2009. Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models. Global Change Biology **15**, 976–991.

Khan HA, Nakamura Y, Furbank RT, Evans JR. 2021. Effect of leaf temperature on the estimation of photosynthetic and other traits of wheat leaves from hyperspectral reflectance. Journal of Experimental Botany **72**, 1271–1281.

Kleinebecker T, Schmidt SR, Fritz C, Smolders AJP, Hölzel N. 2009. Prediction of δ¹³C and δ¹⁵N in plant tissues with near-infrared reflectance spectroscopy. New Phytologist **184**, 732–739.

Kokaly R. 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. Remote Sensing of Environment **67**, 267–287.

Kokaly RF. 2001. Investigating a physical basis for spectroscopic estimates of leaf nitrogen concentration. Remote Sensing of Environment. **75**, 153–161.

Kokaly RF, Asner GP, Ollinger SV, Martin ME, Wessman CA. 2009. Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. Remote Sensing of Environment **114**, S78–S91.

Krishnan A, Williams LJ, McIntosh AR, Abdi H. 2011. Partial least squares (PLS) methods for neuroimaging: a tutorial and review. NeuroImage **56**, 455–475.

Kumar L, Schmidt K, Dury S, Skidmore A. 2002. Imaging spectrometry and vegetation science. In: Meer FD, Jong SMD, eds. Imaging spectrometry. Dordrecht: Springer, 111–155.

Kumarathunge DP, Medlyn BE, Drake JE, *et al.* 2019. Acclimation and adaptation components of the temperature dependence of plant photosynthesis at the global scale. New Phytologist **222**, 768–784.

Martens H, Martens M. 2000. Modified jack-knife estimation of parameter uncertainty in bilinear modelling by Partial Least Squares Regression (PLSR). Food Quality and Preference **11**, 5–16.

Martin ME, Aber JD. 1997. High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. Ecological Applications **7**, 431–443.

Meacham-Hensold K, Montes CM, Wu J, *et al.* 2019. High-throughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity. Remote Sensing of Environment **231**, 111176.

Merzlyak MN, Gitelson AA, Chivkunova OB, Solovchenko AE, Pogosyan SI. 2003. Application of reflectance spectroscopy for analysis of higher plant pigments. Russian Journal of Plant Physiology **50**, 704–710.

Mevik B-H, Wehrens R, Hovde Liland K. 2015. Partial least squares and principal component regression. http://mevik.net/work/software/pls.html.

Myers-Smith IH, Thomas HJD, Bjorkman AD. 2019. Plant traits inform predictions of tundra responses to global change. New Phytologist **221**, 1742–1748.

Noda HM, Motohka T, Murakami K, Muraoka H, Nasahara KN. 2013. Accurate measurement of optical properties of narrow leaves and conifer needles with a typical integrating sphere and spectroradiometer. Plant, Cell & Environment **36**, 1903–1909.

Nozue K, Maloof JN. 2006. Diurnal regulation of plant growth. Plant, Cell & Environment **29**, 396–408.

Pérez-Harguindeguy N, Díaz S, Garnier E, *et al.* 2013. New handbook for standardised measurement of plant functional traits worldwide. Australian Journal of Botany **61**, 167.

Pick TR, Bräutigam A, Schlüter U, *et al.* 2011. Systems analysis of a maize leaf developmental gradient redefines the current C₄ model and provides candidates for regulation. The Plant Cell **23**, 4208–4220.

Polley HW, Yang C, Wilsey BJ, Fay PA. 2020. Spectrally derived values of community leaf dry matter content link shifts in grassland composition with change in biomass production. Remote Sensing in Ecology and Conservation **6**, 344–353.

Reich PB, Walters MB, Ellsworth DS. 1997. From tropics to tundra: global convergence in plant functioning. Proceedings of the National Academy of Sciences, USA **94**, 13730–13734.

Reynolds M, Langridge P. 2016. Physiological breeding. Current Opinion in Plant Biology **31**, 162–171.

Sawatsky ML, Clyde M, Meek F. 2015. Partial least squares regression in the social sciences. The Quantitative Methods for Psychology **11**, 52–62.

Schweiger AK. 2020. Spectral field campaigns: planning and data collection. In: Cavender-Bares J, Gamon JA, Townsend PA, eds. Remote sensing of plant biodiversity. Cham: Springer, 385–423.

Schweiger AK, Cavender-Bares J, Townsend PA, Hobbie SE, Madritch MD, Wang R, Tilman D, Gamon JA. 2018. Plant spectral diversity integrates functional and phylogenetic components of biodiversity and predicts ecosystem function. Nature Ecology & Evolution **2**, 976–982.

Serbin SP. 2012. Spectroscopic determination of leaf nutritional, morphological, and metabolic traits. PhD Dissertation, University of Wisconsin-Madison, Madison, WI.

Serbin SP, Dillaway DN, Kruger EL, Townsend PA. 2012. Leaf optical properties reflect variation in photosynthetic metabolism and its sensitivity to temperature. Journal of Experimental Botany **63**, 489–502.

Serbin SP, Singh A, McNeil BE, Kingdon CC, Townsend PA. 2014. Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. Ecological Applications **24**, 1651–1669.

Serbin SP, Townsend PA. 2020. Scaling functional traits from leaves to canopies. In: Cavender-Bares J, Gamon JA, Townsend PA, eds. Remote sensing of plant biodiversity. Cham: Springer, 43–82.

Serbin SP, Wu J, Ely KS, Kruger EL, Townsend PA, Meng R, Wolfe BT, Chlus A, Wang Z, Rogers A. 2019. From the Arctic to the tropics: multibiome prediction of leaf mass per area using leaf reflectance. New Phytologist **224**, 1557–1568.

Shmueli G, Ray S, Velasquez Estrada JM, Chatla SB. 2016. The elephant in the room: Predictive performance of PLS models. Journal of Business Research **69**, 4552–4564.

Sicher RC, Kremer DF, Harris WG. 1984. Diurnal carbohydrate metabolism of barley primary leaves. Plant Physiology **76**, 165–169.

Silva-Perez V, Molero G, Serbin SP, Condon AG, Reynolds MP, Furbank RT, Evans JR. 2018. Hyperspectral reflectance as a tool to measure biochemical and physiological traits in wheat. Journal of Experimental Botany **69**, 483–496.

Singh A, Serbin SP, McNeil BE, Kingdon CC, Townsend PA. 2015. Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. Ecological Applications **25**, 2180–2197.

Smith NG, Keenan TF, Colin Prentice I, *et al.* 2019. Global photosynthetic capacity is optimized to the environment. Ecology Letters **22**, 506–517.

Thomas HJD, Bjorkman AD, Myers-Smith IH, *et al.* 2020. Global plant trait relationships extend to the climatic extremes of the tundra biome. Nature Communications **11**, 1351.

Townsend PA, Foster JR, Chastain RA, Currie WS. 2003. Application of imaging spectroscopy to mapping canopy nitrogen in the forests of the central Appalachian mountains using Hyperion and AVIRIS. IEEE Transactions on Geoscience and Remote Sensing **41**, 1347–1354

Ustin SL, Roberts DA, Pinzón J, Jacquemoud S, Gardner M, Scheer G, Castañeda CM, Palacios-Orueta A. 1998. Estimating canopy water content of chaparral shrubs using optical methods. Remote Sensing of Environment **65**, 280–291.

Villa P, Bolpagni R, Pinardi M, Tóth VR. 2020. Leaf reflectance can surrogate foliar economics better than physiological traits across macrophyte species. bioRix https://doi.org/10.1101/2020.06.03.131375. [Preprint].

Violle C, Navas M, Vile D, Kazakou E, Fortunel C, Hummel I, Garnier E. 2007. Let the concept of trait be functional! Oikos **116**, 882–892.

Walker AP, Beckerman AP, Gu L, Kattge J, Cernusak LA, Domingues TF, Scales JC, Wohlfahrt G, Wullschleger SD, Woodward FI. 2014. The relationship of leaf photosynthetic traits—V cmax and J max—to leaf nitrogen, leaf phosphorus, and specific leaf area: a meta-analysis and modeling study. Ecology and Evolution **4**, 3218–3235.

Wang Z, Chlus A, Geygan R, Ye Z, Zheng T, Singh A, Couture JJ, Cavender-Bares J, Kruger EL, Townsend PA. 2020. Foliar functional traits from imaging spectroscopy across biomes in eastern North America. New Phytologist **228**, 494–511.

Wold S, Sjöström M, Eriksson L. 2001. PLS-Regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems **58**, 109–130.

Wolter P, Townsend P, Sturtevant B, Kingdon C. 2008. Remote sensing of the distribution and abundance of host species for spruce budworm in Northern Minnesota and Ontario. Remote Sensing of Environment **112**, 3971–3982.

Woodward FI, Diament AD. 1991. Functional approaches to predicting the ecological effects of global change. Functional Ecology **5**, 202–212.

Wright IJ, Reich PB, Westoby M, et al. 2004. The worldwide leaf economics spectrum. Nature **428**, 821–827.

Wu J, Chavana-Bryant C, Prohaska N, et al. 2017. Convergence in relationships between leaf traits, spectra and age across diverse canopy environments and two contrasting tropical forests. New Phytologist **214**, 1033–1048.

Wu J, Rogers A, Albert LP, Ely K, Prohaska N, Wolfe BT, Oliveira RC, Saleska SR, Serbin SP. 2019. Leaf reflectance spectroscopy captures variation in carboxylation capacity across species, canopy environment and leaf age in lowland moist tropical forests. New Phytologist **224**, 663–674.

Yang X, Tang J, Mustard JF, Wu J, Zhao K, Serbin S, Lee J. 2016. Seasonal variability of multiple leaf traits captured by leaf spectroscopy at two temperate deciduous forests. Remote Sensing of Environment **179**, 1–12.

Yendrek CR, Tomaz T, Montes CM, Cao Y, Morse AM, Brown PJ, McIntyre LM, Leakey AD, Ainsworth EA. 2017. High-throughput phenotyping of maize leaf physiological and biochemical traits using hyperspectral reflectance. Plant Physiology **173**, 614–626.

Yoder BJ, Pettigrew-Crosby RE. 1995. Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400–2500 nm) at leaf and canopy scales. Remote Sensing of Environment **53**, 199–211.